Catena 75 (2008) 129–137

Contents lists available at ScienceDirect

Catena

journal homepage: www.elsevier.com/locate/catena

Testing the relationship between instantaneous peak flow and mean daily flow in a Mediterranean Area Southeast Spain

E.V. Taguas [a,*,1], J.L. Ayuso [a,1], A. Pena [a,1], Y. Yuan [b,2], M.C. Sanchez [a,1], J.V. Giraldez [c,3], R. Pérez [a,1]

[a] Department of Rural Engineering, University of Cordoba, Apdo 3048, 14080, Cordoba, Spain
[b] USDA-ARS-NSL, P. O. Box 1157, Oxford, 38655, Mississippi, USA
[c] Department of Agronomy, University of Cordoba, Apdo. 3048, 14080, Cordoba, Spain

ARTICLE INFO

Article history:
Received 13 July 2007
Received in revised form 7 April 2008
Accepted 15 April 2008

Keywords:
Maximum annual instantaneous stream flow
Maximum annual mean daily stream flow
Principal Components Analysis
Semi-arid areas
Spain

ABSTRACT

Extreme hydrologic events have great importance in Southeastern Spain regarding the personal and economic damage they imply. The commonly-used design parameter for hydraulic structures is the maximum annual instantaneous stream flow recorded in conventional gauging stations. However, the majority of available data in Southeastern Spain is mean daily stream flows. This paper explores possible linear relationships between annual instantaneous peak discharge (IPF) and the corresponding (MDF) mean daily stream flow. This relationship was previously explored by other authors such as Fuller [Fuller, W.E., 1914. Flood flows. Trans. Am. Soc. Civ. Eng., 77: 564–617]. Non-linear responses of IPF-–MDF were observed in several study basins. The use of Principal Components Analysis (PCA) allowed characterizing the most important topographic and hydrological attributes of the basins and provided important information about variables that should be included in IPF–MDF regional equations. The key factor to justify the different IPF–MDF relationships in a relatively small area is the nature of the extreme events and their effects on semi-arid soil conditions. In addition, a regional equation to estimate IPF from MDF was developed. This equation was applied to a series of flow of nine stations of the Southeast Basin of Spain, and a significant improvement was achieved when applying this formula in comparison to the traditional method of Fuller. This study indicates possible restrictions to take into account when traditional hydrological model are applied in semi-arid areas.
© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The meteorological conditions of Southeastern Spain include the occurrence of heavy storms that cause devastating floods. The high concentration of population in the coastal tourist resorts of the region and the ephemeral nature of some of the riverbeds require the use of accurate hydrological tools to prevent these storms from causing serious damage. The design of hydraulic structures to control floods is often based on the instantaneous peak flow (IPF) because this may be considerably different to mean flow values, especially in the case of small basins (Fill and Steiner, 2003).

However, the most common hydrological variable recorded by government agencies is the data relating to mean daily flow (MDF). In many cases, use of MDF data for structure design may cause the control structure to be underestimated, with the consequent risk of possible failure.

The relationships between IPF and its corresponding MDF have been well-studied. One of the most commonly accepted methods is based on Fuller (1914) (Eq. (1)), where the IPF and the MDF are related by characteristics of the basin (Gray, 1973; Correia, 1983; Silva and Tucci, 1998).

$$IPF = MDF^{*}\left(1 + 2.66A^{-0.3}\right) \tag{1}$$

where IPF=instantaneous peak flow (m^3/s); MDF=mean daily flow (m^3/s); A=drainage area (km^2).

However, the mechanisms of infiltration and runoff in semi-arid regions, such as in Mediterranean Area, differ from traditionally accepted models applied in humid basins (Nicolau et al., 1996; Calvo-Cases et al., 2003). Therefore, the objectives of this study are: 1) to explore the possible relationships of IPF–MDFIPF-MDF; 2) to study the hydrological and geomorphologic factors that restrict the responses of basins using Principal Components Analysis; 3) to develop regional equations to calculate the peak flow, taking into account the influence of different characteristics of the basins.

2. Materials and methods

2.1. Data used

The study site is situated in an area whose administrative name is the South Mediterranean Basin of Spain. It is comprised of 19

* Corresponding author.
E-mail addresses: ir2tarue@uco.es (E.V. Taguas), ir1aymuj@uco.es (J.L. Ayuso), ir1peala@uco.es (A. Pena), YongPing.Yuan@ARS.USDA.GOV (Y. Yuan), mcsancheztrigo@juntadeandalucia.es (M.C. Sanchez), ag1gicej@uco.es (J.V. Giraldez), ir1pealr@uco.es (R. Pérez).
[1] Tel.: +34 957 218571; fax: +34 957 218563.
[2] Tel.: +662 232 2913; fax: +662 281 5706.
[3] Tel.: +34 957 499008; fax: +34 957 218563.

0341-8162/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.catena.2008.04.015

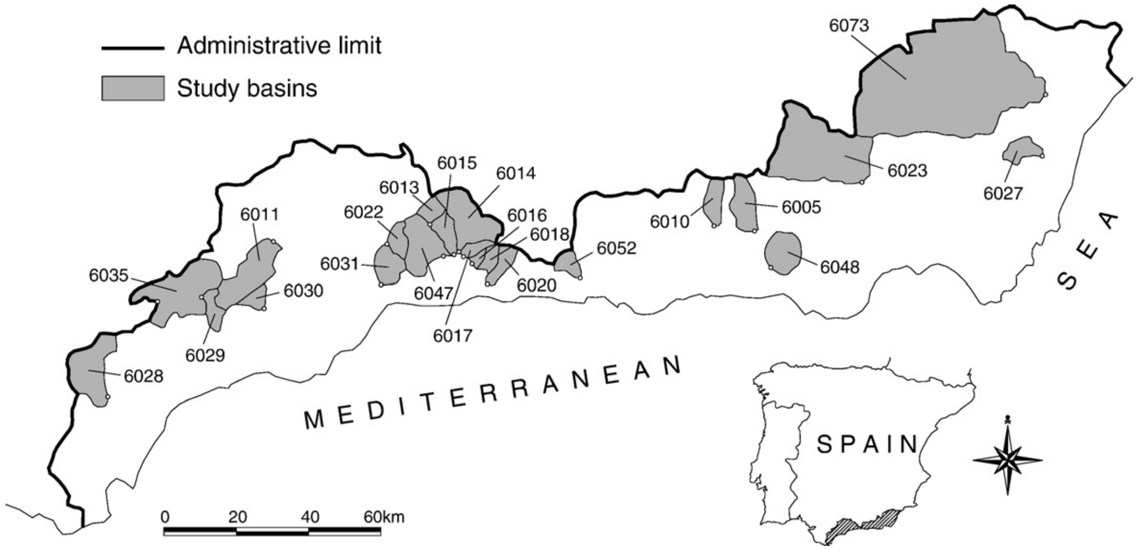

Fig. 1. Location of the stream flow gauging stations (codes) in Southeast Spain.

independent basins situated in a coastal strip about 300 km long. All these basins drain into the Mediterranean Sea in Andalusia (Fig. 1).

The steep topography of the Penibetica mountain ranges, which includes the highest peaks of Spain in Sierra Nevada, causes pronounced climate variations and a wide variation in precipitation regimes within a small area. While the rainfall is considerable in the west, reaching values of over 2000 mm in several places, in the sub-desert conditions of Almeria in the east, as little as 200 mm can be recorded.

On the other hand, over 50% of the surface area includes slopes of over 25%. The steep slopes and the intense heat of summer cause the formation of shallow discontinuous soils of coarse texture. There are some highly appreciated native plants, especially in the Sierra Nevada and Ronda mountain ranges, despite the fact that the plant cover may be scarce on the steep, rocky hillsides with poor soil.

The population in this area is around 2,100,000, but in summer it can rise to 2,700,000 due to the influx of tourists from all over Europe. The concentration of the population on the coastal perimeter is due to the fact that tourism and agriculture are the main economic activities (Agencia Andaluza del Agua, 2005).

Mean daily flow and peak flow data from 23 stations were collected from the Confederacion Hidrografica del Sur (Southern Hydrographical Confederation) website (Table 1), corresponding to the time period from 1960 to 1995. They were checked for consistency and one station was removed because some peak flow values were less than the corresponding mean daily flow.

Drainage area, stream slope, minimum and maximum elevation and length of the stream were obtained from the Digital Elevation Model of Andalusia (ICA, 1999) with the Geographical Information System (GIS) Arcview 3.2. (ESRI, 2000) for the 22 sub-basins (Table 1). The soil data and land use distribution (Table 2) were extracted with Arcview 3.2. (ESRI, 2000) from the Interactive Atlas of Andalusia (ICA, 2000).

Mean annual rainfall values for the sub-basins were calculated with the method of Thiessen polygons by using data collected from rainfall gauge stations from the "Confederacion Hidrografica del Sur" (Southern Hydrographical Confederation) website (Table 2). Then, the data for mean daily flow in 20 stations was used to analyze the hydrological regime.

Table 1
Stream flow gauging stations and some attributes of the drainage basins

Code	Name	Data (n)	Data period	X UTM (m)	Y UTM (m)	Drainage area (km^2)	Stream slope (%)	Min elevation (outlet) (m)	Max elevation (m)	Stream length (m)
6005	Tosquillas	8	1960–89	496,285	4,087,850	120	8.8	420	2781	21,033
6010	Narila	13	1971–91	483,280	4,090,240	67	11.4	960	2905	16,296
6011	Ardales	14	1961–93	335,335	4,084,145	211	3.3	346	1919	31,446
6013	Alfartanejo	15	1961–91	388,415	4,091,355	39	3.6	810	1477	11,123
6014	Cortijo	13	1960–85	396,595	4,080,310	119	3.8	160	1497	27,455
6015	Vinuela	16	1960–92	398,055	4,080,355	67	6.6	150	2065	25,820
6016	Gonzalez	13	1960–91	400,495	4,081,345	13	19.1	280	2065	8704
6017	Pasada	15	1965–91	401,330	4,078,715	12	13.7	160	2065	10,811
6018	Hoya	14	1965–91	401,855	4,077,165	47	7.6	180	2065	15,889
6020	Umbria	9	1961–91	407,745	4,071,800	67	7.0	120	1788	15,558
6022	Casabermeja	9	1978–91	373,720	4,084,535	65	3.0	500	1039	11,044
6023	Chono	17	1960–92	532,050	4,103,470	616	2.4	512	2519	50,149
6027	Alfaix	11	1965–91	592,215	4,111,310	68	2.5	115	887	15,330
6028	Jimena	8	1968–93	280,185	4,034,410	245	1.9	29	1082	30,042
6029	Molino	8	1973–93	310,435	4,067,085	66	3.1	708	1788	15,249
6030	Cueva	21	1960–93	300,135	4,067,165	51	2.5	430	1545	24,707
6031	Agujero	9	1968–83	372,130	4,070,935	153	2.2	80	1031	33,356
6035	Millanas	12	1969–91	331,960	4,063,770	38	7.2	215	1600	7928
6047	Salto	15	1969–91	392,585	4,078,820	182	3.3	139	1481	22,599
6048	Ventilla	10	1971–89	502,055	4,075,830	143	7.0	200	2220	20,844
6052	Cazulas	6	1969–86	437,385	4,074,305	43	7.6	290	1717	11,398
6073	Barbara	5	1980–89	591,575	4,134,115	1850	1.0	175	2168	77,567

Table 2
Mean annual rainfall, predominant soil texture and land uses of the basins

Code	Name	Annual mean rainfall (mm)	Predominant soil texture (%)	Predominant land use (%)
6005	Tosquillas	402	Loams or sandy loams	Rangeland (59.5%)–Cropland (21.7%)
6010	Narila	608	Loams or sandy loams	Rangeland (45.2%)–Urban (35.7%)
6011	Ardales	596	Clayey	Rangeland (63.4%)–Cropland (35.5%)
6013	Alfartanejo	888	Clayey	Cropland (42.8%)–Rangeland (39.4%)
6014	Cortijo	620	Clayey	Cropland (55.7%)–Rangeland (36.1%)
6015	Vinuela	705	Loamy limestone, clayey	Rangeland (65.9%)–Cropland (30.4%)
6016	Gonzalez	489	Clayey, abundant rocky limestone	Cropland (82.3%)–Rangeland (17.6%)
6017	Pasada	445	Loams or sandy loams	Cropland (53.8%)–Rangeland (42.4%)
6018	Hoya	497	Loams or sandy loams	Cropland (76.8%)–Rangeland (22.2%)
6020	Umbria	650	Loams or sandy loams	Rangeland (49.8%)–Cropland 30.1%
6022	Casabermeja	612	Loamy/clayey	Rangeland (47.9%)–Cropland (41.0%)
6023	Chono	271	Loams or sandy loams	Cropland (62.4%)–Rangeland (34.9%)
6027	Alfaix	336	Loams or sandy loams	Rangeland (76.9%)–Cropland (11.0%)
6028	Jimena	1024	Sandy	Rangeland (50.1%)–Urban (41.2%)
6029	Molino	805	Clayey, abundant rocky limestone	Rangeland (74.3%)–Cropland (19.2%)
6030	Cueva	722	Loamy limestone	Rangeland (58.7%)–Cropland (33.3%)
6031	Agujero	622	Loams or sandy loams	Cropland (56.8%)–Rangeland (40.3%)
6035	Millanas	687	Clayey/loamy	Rangeland (71.1%)–Cropland (20.1%)
6047	Salto	631	Loamy/clayey	Cropland (83.5%)
6048	Ventilla	354	Volcanic rock	Cropland (66.4%)–Urban (22.1%)
6052	Cazulas	612	Clayey, abundant rocky limestone	Rangeland (71.2%)
6073	Barbara	309	Loamy/clayey	Cropland (46.8%)–Rangeland (40.6%)

2.2. Preliminary data analysis: graphic study and linear regression fits

As the first step of the study, a correlation analysis between the maximum annual instantaneous IPF and mean daily maximum MDF flow was carried out. The data was analyzed through graphical study and linear regressions of IPF = a * MDF were fitted to the data, based on the relationship obtained by Fuller (1914). In addition, an analysis of variance was applied to the regressions (F-Snedecor test) to examine if the relationship between dependent and independent variables occurs randomly for a significance level of 0.05. The statistical F (Eq. (2)) was used (Steel and Torrie, 1980, Chap. 10).

$$F = \frac{\frac{\left[\Sigma\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)\right]^2}{\Sigma\left(X-\overline{X}\right)^2}}{\Sigma\left(Y-\overline{Y}\right)^2-\frac{\left[\Sigma\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)\right]^2}{\Sigma\left(X-\overline{X}\right)^2}} \tag{2}$$

where X are MDF data and Y are IPF data. This value F was compared to F* for a significance level of 0.05; the number of independent variables (k) is 1, and the degree of freedom is $n-2$ (n = number of pairs of data).

Different types of fits were classified according to different tendencies observed: good linear fits ($0.75<R^2<1$), acceptable linear fits ($0.55<R^2<0.75$), non-linear fits (due to low R^2 or failed F-Snedecor test) and the group without representative samples for the fits (due to low number of data, improvements of R^2 by extreme values or failed F-Snedecor test). In addition, a Student's t test was applied to contrast the homogeneity of the regression coefficients "a" in basins with good and acceptable linear fits (Eq. (3)), determining whether or not the expressions obtained produced estimates in the same confidence level (Steel and Torrie, 1980, Chap. 10):

$$t = \frac{a_1-a_2}{\sqrt{S_p^2\left[1/\Sigma\left(X_{1j}-\overline{X}_1\right)^2+1/\Sigma\left(X_{2j}-\overline{X}_2\right)^2\right]}} \tag{3}$$

where a_1 and a_2 are the coefficients of the regression and $\Sigma(X_{1j}-\overline{X}_1)^2 y$ $\Sigma(X_{2j}-\overline{X}_2)^2$ is the sum of squares of the independent variable of the first and second stream flow gauging station selected. The t statistic is distributed as a Student's t with n_1+n_2-4 degrees of freedom, where n_1 and n_2 are the number of data from each station. The term s_p^2 is the best estimate of the variation with respect to the regression. This is considered to be the combined sums of the residual squares of the two regressions divided by the combined degrees of freedom.

Finally, global expressions IPF–MDF were obtained, calculating the regression coefficients corresponding to the data of the groups observed.

2.3. Exploring the relationships IPF–MDF in the study basins: Principal Components Analysis

Equations derived by various authors (Gray, 1973; Correia, 1983; Tucci, 1991) based on Fuller's studies (Eq. (1)) have set the drainage area as the only independent variable to calculate IPF. However, new variables could be included in the equations provided that they improve the results. The aim of this section is to study the existence of groups of similar basins where these equations would be useful as well as the restrictions in the relationships IPF–MDF.

The statistical method of Principal Components Analysis (PCA) was used to determine the factors that make a group of individuals (basins) different or similar. The objective of this method was to isolate the basins by means of linearly-grouped common features (Malinowski, 1991, Chap. 1.2.). PCA changes the initial variables, all of which are more or less correlated, into new, non-correlated synthetic variables called principal components, increasing the efficiency of the analysis. The variables considered in the analysis should model the phenomenon studied in the closest possible way (Philippeau, 1986).

We have considered a set of measures $\{X_i(w_j)\}$ of p variables, $\{X_i\}$ on a set of n subjects $\{w_j\}$, which correspond to the basins studied. This data allows us to define a matrix with a dimension of $p\times n$, or an initial data matrix. The following steps are taken: 1) to calculate the correlations matrix (Eq. (4)); 2) to make the correlations matrix diagonal $\{\rho_{i,i-1}\}$ and obtain the eigenvalues λ_α, and eigenvectors u_α, (Eqs. (5) and (6)); 3) to calculate the projections of individual cases on the main axis ($\left\{w_j^*\right\}^{(p)}$, (Eq. 7)); 4) to calculate the projection of the variables on the principal plane ($\rho_{ip}^*(\{w_j^*\}^{(p)}$, Eqs. 8) and 5) to test the quality of the representation of the variables and individuals to verify the quality of the representation in the principal plane.

$$\{\rho_{i,i-1}\} = 1/n\sum_{i=1}^{n}\ \{[X_i(w_j)-\mu_i]/\sigma_i\}\{[X_{i-1}(w_j)-\mu_{i-1}]/\sigma_{i-1}\} \tag{4}$$

$$Det[\{\rho_{i,i-1}\}-\lambda_\alpha I] = 0 \tag{5}$$

$$[\{\rho_{i,i-1}\}-\lambda_a I]\ x\ \mu_\alpha, = 0 \tag{6}$$

$$\{Z_i(w_j)\}\ x\ \mu_{\alpha.,} = \left\{w_j^*\right\}^{(p)} \tag{7}$$

$$\rho_{ip}^*\left(\left\{w_j^*\right\}^{(p)}, \{Z_i(w_j)\}\right) \tag{8}$$

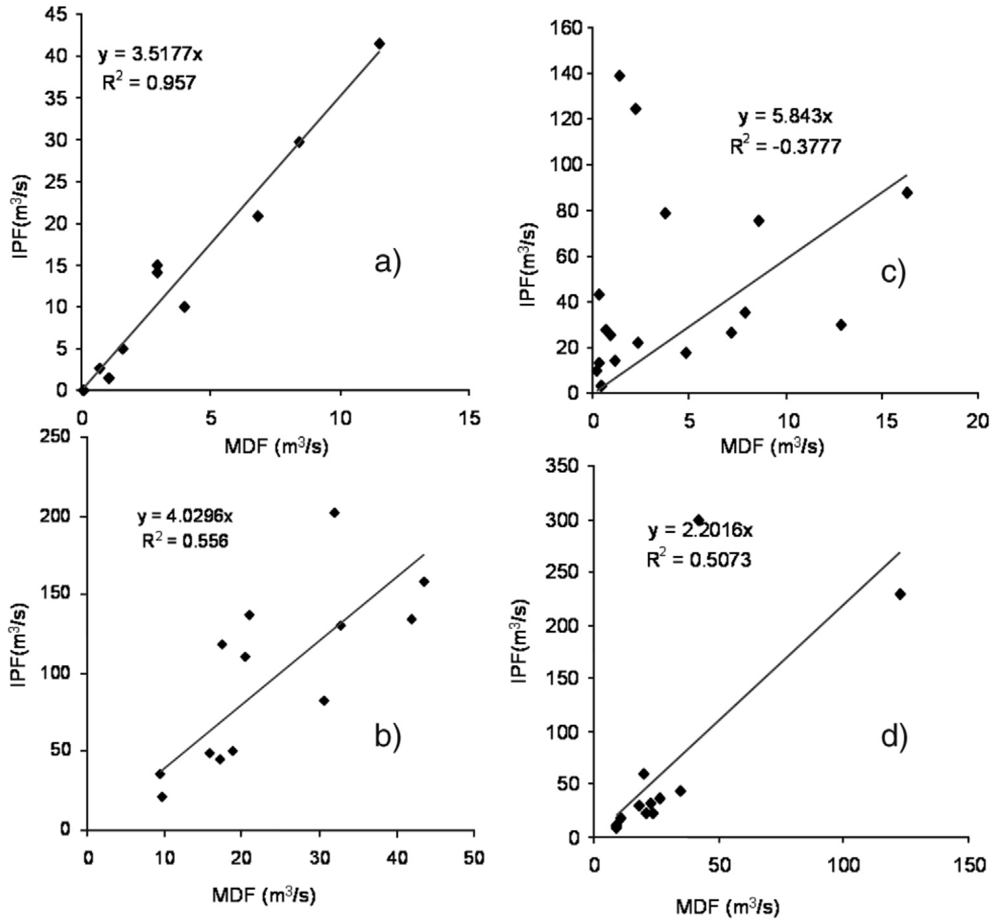

Fig. 2. Representative fits of maximum annual instantaneous peak flow IPF (m^3/s) and the corresponding maximum daily mean flow MDF (m^3/s) for the stations: a) 6029—Casabermeja; b) 6047—Salto; c) 6023—Chono; d) 6021—Ardales.

The following variables were included:

- 24 H rain depth for the 100-year event (P, T=100). This parameter is an indicator of the extreme rainfall events regime in each basin. Despite the high level of uncertainty because the majority of data series are short or incomplete (between 30 and 50 years), the 100-year return period was chosen to emphasize possible differences between nearby basins and to select a representative frequency value for the rainfall distribution.
- Mean annual rainfall in the basin (P). This variable shows the decreasing trend in rainfall in the east of the region. We applied the method of Thiessen polygons with rainfall data of meteorological stations in the basins using a tool of Arcview 3.2. (ESRI, 2000).
- Linear moments or L-moments of mean daily flows in each basin, computed for twenty of the stations, (with the exception of the Cortijo and Agujero stations due to incomplete and short series) for the period 1985–2000. The L-moments are an alternative to the traditional method (conventional moments) for the distribution functions of the extreme events, thus permitting their comparison once they are normalized (Álvarez et al., 1999). These are similar to ordinary moments as they provide measures of location, dispersion, skewness, and kurtosis, but are calculated from the linear combinations of data (Hosking and Wallis, 1997). Because the statistical parameters estimated are not over-sensitive to extreme values, they are very useful when working with maximum mean daily flows subject to large annual variations. The mean, the coefficient of variation (CLV), the skewness coefficient (CLS) and the kurtosis coefficient (CLK) of the mean daily flows were used in order to define the type of hydrological regime of each basin.
- Maximum–minimum instantaneous flow rate ratio (IPF_{max}/IPF_{min}). This variable is an indicator of irregularity as it expresses the inter-annual variability between the maximum flows of wet and dry years in each basin.

Table 3
Stations with coefficients "a" derived from linear fit (IPF=a*MDF), with their coefficients of determination (R^2) and groups obtained from homogeneity test of "a"

Code	Name	a	R^2	Group (test of homogeneity B)
6031	Agujero	3.26	0.76	Group 1
6013	Alfartanejo	4.22	0.76	
6029	Casabermeja	3.52	0.96	
6021	Cortijo	3.79	0.75	
6017	Narila	2.91	0.93	
6015	Vinuela	4.77	0.78	
6020	Umbria	3.07	0.78	
6035	Millanas	1.64	0.96	Group 2
6030	Cueva	1.03	0.99	
6033	Molino	2.51	0.64	–
6028	Jimena	2.62	0.56	–
6047	Salto	4.03	0.56	–
6018	Hoya			Non-linear relationship IPF–MDF or excluded to the analysis
6025	Gonzalez*			
6048	Ventilla*			
6005	Tosquillas*	–	–	
6052	Cazulas*			
6023	Chono*			
6017	Pasada*			
6027	Alfaix*			
6073	Barbara			
6011	Ardales			

* Indicates the basins that did not pass the F-Snedecor test (level of significance=0.05).

– The physical factors selected for the study include the *urbanized or uncultivated/rangeland sections* of the basins.
– Finally, the values corresponding to basin area (A), slope (S) and length (L) of the main channel were included in the analysis. These factors are related to the time of concentration and general shape of hydrographs.

In the environmental study of basins, numerous geomorphologic features were taken into account to explain different IPF–MDF behaviour in the basins. However, because the qualitative characteristics (parent material, soil type, climate regimes, etc) are difficult to be included either into PCA or into equations, they were not considered. Numerical variables (such as annual runoff coefficient, elevation data, texture, etc.) that reached analogue average values and the supporting information were not significant and were also not considered. The variables defined comprise only a fraction of those that could be included in the PCA. However, they provide the most relevant information about the hydrological regime and the long-term responses of the basins.

Principal Components Analysis was performed on the matrix of the *standardized* attributes defined above, using 20 of the 22 basins studied — we used Statistica 6.0 (StatSoft, 2001) software. All the individuals and variables were assigned the same weight and actively intervened in the analysis. The quality of the individual basins was tested by squaring the coordinates of their matrix on the vector space of the principal axes which are also the director cosines of the central-individual projection and indicators of the proximity of the axes. The sum of the values for the first two components is considered another index of quality. Values lower than 0.20 indicate that the projection of the individual is not adequate and cannot be compared to the other individuals because the point is located far from the plane (Philippeau, 1986). Finally, using the new coordinates of the individuals calculated by PCA, the situation on the principal plane was evaluated to judge trends in similar groups of basins.

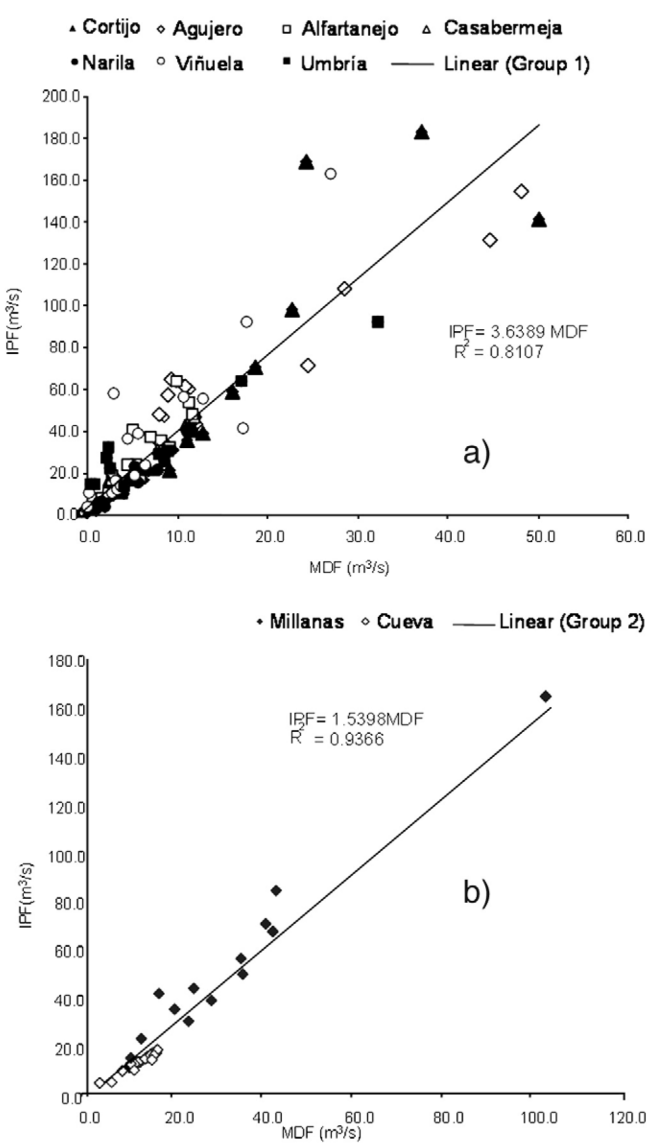

Fig. 3. Linear regressions estimated in basins showing homogeneous regressions: a) Group 1 and b) Group 2.

Table 4
Eigenvalues on the three principal axes and percentage of total variance that they explain

Main axis	Eigenvalues	Variance	Accumulated variance (%)
Axis 1	3.48	0.29	29.0
Axis 2	2.91	0.24	53.0
Axis 3	2.38	0.20	73.0

2.4. Adjustment of an IPF–MDF regional equation

In this case, we grouped the basins with good linear fits to adjust an equation that takes into account the information provided by the PCA and can be directly used or easily adjusted for different combinations of data in the basins of the Mediterranean Area Southeast Spain.

Fuller's expression is very easy to apply and it is useful for the basins where linear relations are observed. On the other hand, Silva and Tucci (1998) used a classical log-linear adjustment to obtain the IPF/MDF rate according to Eq. (9):

$$\mathrm{IPF/MDF} = a^{*}A^{b}L^{c}D^{d}T^{e} \tag{9}$$

where a = regression constant; b, c, d and e = regressions coefficients and A, L, D and T = physiographic characteristics of the basins. Two similar expressions were adjusted by the least squares model (Statistica 6.0; StatSoft, 2001) and their results are compared to values derived by Fuller's equation.

Finally, the comparison of predictions takes into account bias and RMSE, computed using relative values (Eqs. (10) and (11)) due to the wide range of values of flow. This procedure was suggested by several authors (Stedinger, 1980; Fill, 1994; Fill and Steiner, 2003) for the evaluation and comparison of estimator performance.

$$\mathrm{bias} = \frac{\Sigma y_i}{n} - 1 \tag{10}$$

$$\mathrm{RMSE} = \sqrt{\frac{\Sigma(y_i - 1)^2}{n}} \tag{11}$$

where y_i = ratio between the estimated and the observed value of the peak flow and n = total number of events.

3. Results

3.1. Preliminary analysis

Once the linear regression fits were done, four different IPF–MDF tendencies were observed: good fit ($0.75 < R^2 < 1$, Fig. 2a), acceptable fit ($0.55 < R^2 < 0.75$, Fig. 2b) and non-linear fit (Fig. 2c). A group of basins (Fig. 2d) was excluded from the analysis due to either a lack of data or a sample in which the extreme values improved the coefficient of

Table 5
Principal components of axes 1, 2 and 3 and quality index of the principal axes 1 and 2 and axes 1, 2 and 3

Attributes	Axis 1	Axis 2	Axis 3	Quality index (axes 1 and 2)	Quality index (axes 1, 2 and 3)
P, *T* = 100 (mm)	−0.101	0.845	0.091	0.724	0.732
% Rangeland	0.261	0.639	0.026	0.477	0.478
Drainage area (km^2)	0.161	−0.361	0.861	0.156	0.898
Slope (%)	−0.301	−0.358	−0.678	0.219	0.679
Length channel (m)	0.150	−0.356	0.896	0.149	0.953
Annual *P* (mm)	−0.506	0.807	0.022	0.907	0.908
Mean MDF	−0.294	0.566	0.513	0.407	0.671
CLV MDF	0.388	0.628	0.052	0.545	0.548
CLS MDF	0.951	0.088	−0.186	0.911	0.946
CLK MDF	0.944	0.054	−0.147	0.894	0.916
IPF_{max}/IPF_{min}	0.949	0.143	−0.072	0.922	0.927
% Urban use	−0.094	0.193	0.132	0.046	0.064

determination (R^2) of the regressions or due to the fact that the relationship between the variables occurred randomly for a significance level of 0.05 (*F*-Snedecor test, Eq. (2)).

Table 3 shows the estimated values of the coefficient (*a*) and the coefficient of determination (R^2) of regression for basins with good and acceptable fits. The last column shows the two groups (1 and 2) derived from the homogeneity test (Eq. (3)). The basins Jimena, Molino and Salto were excluded from the groups 1 and 2 according to the results of the Student's *t* test. The IPF–MDF relationships in Jimena and Molino tend to be different due to the fact they show more humid rainfall and flow regimes (Table 2); in fact, both basins present the lowest regression slopes ("*a*" in Table 3), that explain smaller flow variations in them than in the rest basins. In the case of Salto, although the fit is not very good (R^2 = 0.56) the value of the slope is one of the highest (Table 3). These features can be associated to the role of agricultural land use (84% of the surface, Table 2) and different management operations along the year which influenced soil conditions. Different soil conditions resulted in different basin responses to precipitation. In the end, the group of basins without IPF–MDF linear fits and those which have been separated from the group with good and acceptable fits are presented.

To sum up, the global expressions IPF–MDF (Fig. 3) adjusted to data group 1 and data group 2 are shown in Eqs. (12) and (13). The relative root mean square errors between the observed and calculated data are notably high in both groups (37.6% and 35.5%).

$$IPF = 3.64^{*}MDF \tag{12}$$

$$IPF = 1.54^{*}MDF \tag{13}$$

This large relative error is due to the variety of basins or stations used in the analysis as well as to the contrast between IPF recorded in rainy years as opposed to dry years. The root mean square error is 10.5 m^3/s for group 1 and 6.58 m^3/s for group 2.

3.2. PCA: linear fit versus non-linear fit

The results of the analysis are summarized in Tables 4 and 5 and Fig. 4. The tables show the new coordinates and the quality index of the representation for the first two and three axes, as well as the eigenvalues of each variable with its quality index. Fig. 4 shows the location of the basins in relation to the variables that are correlated to the principal plane (axes 1 and 2) and the groups discriminated according to the regressions.

The first principal axis represents 29% of the variance of the system and is an indicator of the CLS, CLK and IPF_{max}/IPF_{min} variables, which are closely correlated to axis 1 (Table 5). The basins are thus distributed on the principal plane according to the irregularity of their hydrological regime. The basin that displays the most irregular behaviour is the Alfaix basin (first quadrant, Fig. 4), followed by Barbara and Chono. These are the most easterly basins and are characterized by their inter-annual and intra-annual irregularity as shown in the component on the first principal axis. It is interesting to note that mean annual rainfall (*P*) correlates negatively with the above variables, so that more irregularity of MDF could be related to less mean rainfall (Table 5).

The second axis, with 24% variance, correlates more closely with the daily rainfall with a 100-year return period (*P*, *T* = 100), mean annual rainfall (*P*), coefficient of variation of MDF (CLV) and

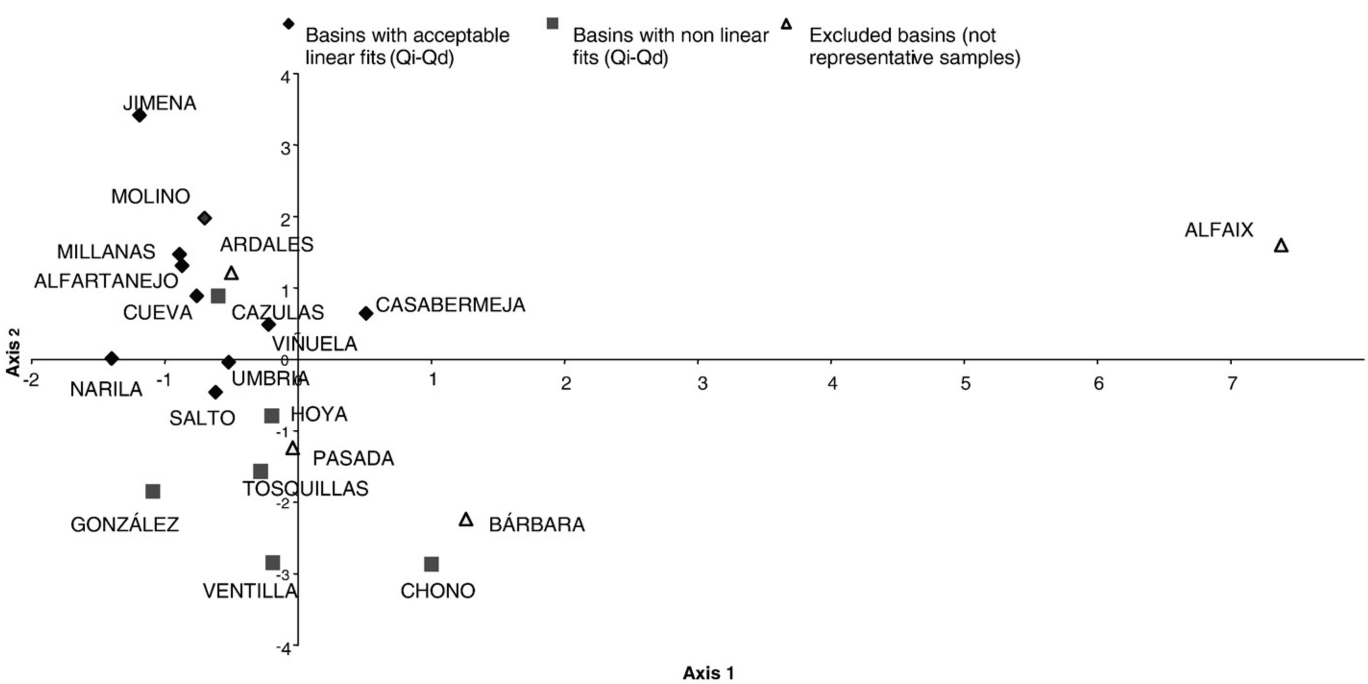

Fig. 4. Representation of the basins using the two principal axes.

Table 6
Comparison of regional equations in Southeast Spain; A (km^2); P (mm); S (%)

Formula	Equation	Parameters				R^2	Bias	RMSE
		a	b	c	d			
Fuller	$IPF=MDF*(1+a*A^b)$	0.0454	0.8169	–	–	0.71	0.051	0.598
Eq. (1)	$IPF=a*MDF*A^b*P^c$	5.91×10^{-4}	0.5723	0.9082	–	0.74	0.069	0.612
Eq. (2)	$IPF=a*MDF*P^b*A^c*S^d$	3.51×10^{-14}	3.9972	1.1575	0.6191	0.82	0.046	0.444

uncultivated or rangeland section (% rangeland). Thus, the basins are distributed on the second principal axis depending on the amount of rainfall they receive, flow variability, and land use. Specifically, vegetation and soil use are key factors in controlling the intensity and frequency of flooding events (Mitchel, 1990).

The third axis (20% of the variance) shows the effect of the physical parameters: *basin area* (A), *slope* (S) *and channel length* (L). As expected, all of these parameters are correlated to the mean flow magnitude, but have no correlation to the rest of the climate components. The large range of drainage areas must be considered because of spatial variability of rainfall in big basins such as Barbara (1850 km^2) for instance, so different mechanism of responses could be observed.

As shown in Fig. 4, according to the distribution on the principal axes, which explained 53% of the variance, four different groups could be observed (two of which are individual). The first of these is the Alfaix basin, which has an extremely variable rainfall regime. On the contrary, the Jimena basin is a very wet area, with over 1000 mm of annual rainfall and a large section of uncultivated land/rangeland that does not resemble the rest of the basins — as demonstrated by the homogeneity test of its flow data.

The remaining basins present similar regimes of MDF but with differences in the rainfall and land use. The rainiest basins are located on the first and second quadrant. As is observed in Fig. 4, the majority of basins with good linear fits are located on the second quadrant with the exception of Casabermeja and Umbria. Despite different locations on the principal plane, the most important issue is the proximity among all basins with acceptable fits and the interpretation of their differences and their similarities related to IPF–MDF behaviour. In the case of Casabermeja, it presents a larger irregularity of flow regime (or variables correlated to the first principal axis: CLS, CLK, IPF_{max}/IPF_{min}, Table 5) in spite of their notable rainfall values. On the contrary, Umbria shows lower values for the linear combination of variables correlated to the second principal axis (especially $P(T=100)$ and CLV) than the rest of the group with good linear fits. Finally, the third and fourth quadrants contain all the basins with the worst fits. To sum up, the group located on the first and second quadrant with acceptable linear fits show the highest values of P, P ($T=100$) and Mean MDF and the lowest of CLS, CLK, and IPF_{max}/IPF_{min}, which are related to wetter rainfall regimes, larger mean daily flows and more regular annual distributions of MDF. The role of the land use (% rangeland) may be interpreted as a cause of the controlling of the basin responses and/or as a consequence of greater amount of rainfall. Thus, these features illustrate the importance of hydrological regularity to justify the linear behaviour IPF–MDF.

3.3. Regional equations to calculate IPF in the Southeast Basin of Spain

The mean annual precipitation (P) has been included in the proposed Eq. (1) (Table 6). In contrast, the variables associated to the irregularity of hydrological regime IPF_{max}/IPF_{min} and to the shape of distribution of frequencies of MDF (Mean MDF, CLS and CLK) have not been included because the equation has been designed to calculate IPF values for basins where this data is not available. On the other hand, it is difficult in most basins of the South of Spain to find complete or long enough annual series of mean daily flows to calculate good estimations of CLS and CLK. Finally, there should be a compromise between number of variables in the equation and its simplicity for the application; especially in comparison with Fuller's equation in use. Despite the information supported by those variables about the IPF–MDF behaviour, their inclusion into the new equation is not competitive.

Thus, adding only the mean annual precipitation P to the equation did not explain the irregularity of hydrological regime. The bias and RMSE of Eq. (1) are worse with regards to Fuller's statistics, despite a better coefficient of determination (R^2). In the case of Fuller's equation, it is observed that most of the deviation (y_i) is concentrated in the IPFs predicted for Cueva (Table 7), which are clearly overestimated, while the bias and the RMSE are lower for IPFs estimated for Agujero, Cortijo and Umbria. In Eq. (1), Cueva shows bigger bias and RMSE ($y_i>2$) than the prediction of Fuller's equation, while similar deviations are observed in the rest basins relative to Fuller's prediction. Therefore, worse bias and RMSE are observed for Eq. (1) as a consequence of the lesser total sums of y_i (Eq. (10)) as well as the larger total sum of $(y_i-1)^2$ in Eq. (11).

On the contrary, the slope of the channel (S) has been used (Eq. (2), Table 6) to introduce a simple topographical component which is easy to evaluate. Table 6 shows the equation, the value of the parameters, the coefficient of determination of the adjustment, the RMSE (Eq. (10)) and the bias (Eq. (11)). It can be observed that the proposed Eq. (2) presents lower RMSE and bias than Fuller's equation and Eq. (1). The coefficient of determination R^2 or proportion of variability that is accounted for by the statistical model is also higher for the proposed equation.

The drainage area and annual rainfall of the 9 basins whose data participated in the development of regional equation ranged from 39 to 182 km^2 and from 608 to 888 mm. Fig. 5 shows the observed and predicted values of IPF. It can be observed that the deviation is higher with greater peak flow values than with lower peak flow values. The RMSE is estimated at 16.0 m^3/s.

4. Discussion

The application of PCA has, firstly, to explain different behaviour between basins both with good linear IPF–MDF fits and without them, and secondly, to find common factors in the basins with linear adjustments that may be included in a regional equation, since similar IPF–MDF linear relationships had been observed in the preliminary analysis. In the heterogeneous group of basins without linear fits, the IPF–MDF pattern is unknown and the PCA is not useful to describe possible

Table 7
Bias and root mean square error (RMSE) with relative values derived from IPF–MDF fits with Fuller's equation, Eqs. (1) and (2)

Basin	Bias (Fuller)	RMSE (Fuller)	Bias (Eq. (1))	RMSE (Eq. (1))	Bias (Eq. (2))	RMSE (Eq. (2))
6031Agujero	−0.05	0.28	−0.08	0.28	−0.28	0.35
6013 Alfartanejo	−0.52	0.54	−0.43	0.45	−0.18	0.26
6022 Casabermeja	−0.09	0.50	−0.08	0.51	−0.30	0.48
6014 Cortijo	0.03	0.27	0.08	0.29	0.36	0.51
6010 Narila	0.18	0.51	0.08	0.45	0.35	0.65
6015 Vinuela	−0.42	0.49	−0.39	0.47	−0.14	0.39
6020 Umbria	−0.42	0.51	−0.44	0.51	−0.36	0.47
6035 Millanas	0.12	0.21	0.06	0.18	0.04	0.17
6030 Cueva	1.05	1.06	1.14	1.14	0.51	0.51

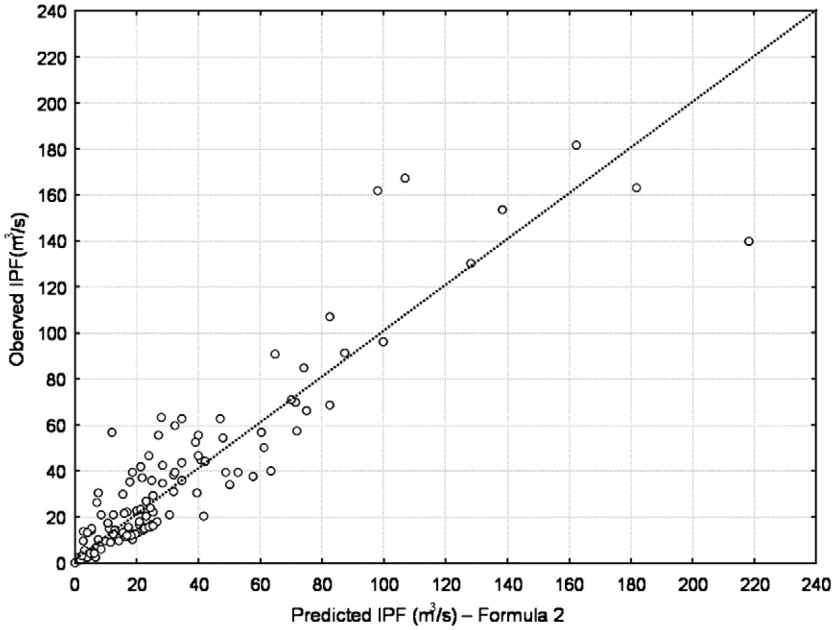

Fig. 5. Observed and predicted IPF values from the proposed regional equation for the data from 9 basins (Agujero, Alfartanejo, Casabermeja, Cortijo, Narila, Vinuela, Umbria, Millanas, and Cueva) with good linear IPF–MDF fits.

adjustment models, only to examine differences in the cases of basins with good adjustments.

The most important question derived from this analysis is why several basins present linear fits of IPF–MDF and in the rest, the regional equation obtained is not valid. In the PCA, we observed the tendency of the basins to be grouped mainly according to climate. On the other hand, the interaction between climate and soil in arid and semi-arid areas is a highly complex matter. Leopold et al. (1964) demonstrated that climate has the largest impact on basins; however, soil conditions will largely determine the response (Cerdá and Lavee, 1995; Kosmas et al., 1997; Cerdá, 1997; Cerdá, 1998). Thus, the range of the precipitation values in the basins with acceptable fits and the characteristics or nature of rainfall could explain the difference of responses in such a small area.

The rainfall regime in the Mediterranean Area is partially determined by large-scale atmospheric circulation patterns such as the location of the Azores anticyclone. This large-scale flow of air interacts with the orography and the area's continental land mass, and local factors (Wigley, 1992) are also involved. Thus, the position of a range of mountains situated in the West and the Center of the Southern Basin could justify some of the differences, since the basins without a linear fit are located mainly in the East (Chono, Tosquillas, Ventilla and Alfaix).

On the other hand, the rainfall is highly seasonal, with only 15% occurring in the summer months. Furthermore, the summer rainfall tends to be highly convective (Mulligan, 1998). In the case of the Southern Basin, 60–70% of the annual rainfall occurs in autumn and winter, although it can also be abundant (as much as 30%) in inland areas during the spring (Landsberg, 1970, Chap. 5). We have observed that in several basins with non-linear fits, most of the maximum peak flows have occurred in summer (ranged from 0–60%, Table 8). This contrasts with basins having good and acceptable linear fits, whose maximum values are reached at other times of the year, mainly in autumn and winter.

Studies such as those carried out by Puigdefábregas et al. (1998) suggest two runoff mechanisms associated to different seasons of the year. These mechanisms may explain the different behaviour of relationships IPF–MDF. The linear fits are associated to a type of surface flow which is the result of the saturation of the first soil horizons, due to the prolonged rainfall which is common in autumn and winter. The conditions of moisture after continued rainfall are relatively homogeneous and similar, especially in basins with areas ranging between 40 and 200 km^2.

On the other hand, the peak flows during summer may be provoked by excess of non-infiltrated flow or Hortonian flow that contributes to runoff according to the intensity of rainfall and the state of the soil (cover, hydraulic properties and previous moisture). Authors such as Calvo-Cases et al. (2003) describe the generation of runoff in hillsides in South-east Spain as a mosaic of sink and source areas of runoff. Under conditions of high evapotranspiration, these areas would have less connectivity and the flow in channels would seldom be affected by the very short, but very intense storms which occur in summer. In contrast, in rainy seasons, the generation of runoff may be more continuous in space and time. The drainage area size

Table 8
Seasonal distribution of the maximum peak flow

Code	Name	Number of data	% Autumn	% Winter	% Spring	% Summer
6005	Tosquillas	8	62.5	12.5	12.5	12.5
6010	Narila	13	53.9	46.2	0.0	0.0
6011	Ardales	14	14.3	78.6	0.0	7.1
6013	Alfartanejo	15	26.7	66.7	6.7	0.0
6014	Cortijo	13	38.5	53.9	7.7	0.0
6015	Vinuela	16	43.8	50.0	6.3	0.0
6016	Gonzalez	13	53.9	46.2	0.0	0.0
6017	Pasada	15	46.7	40.0	6.7	6.7
6018	Hoya	14	50.0	42.9	7.1	0.0
6020	Umbria	9	66.7	33.3	0.0	0.0
6022	Casabermeja	9	55.6	44.4	0.0	0.0
6023	Chono	17	41.2	17.7	23.5	17.7
6027	Alfaix	11	54.6	36.4	9.1	0.0
6028	Jimena	8	50.0	50.0	0.0	0.0
6029	Molino	8	50.0	50.0	0.0	0.0
6030	Cueva	21	19.1	66.7	14.3	0.0
6031	Agujero	9	44.4	33.3	22.2	0.0
6035	Millanas	12	58.3	41.7	0.0	0.0
6047	Salto	15	40.0	53.3	6.7	0.0
6048	Ventilla	10	50.0	30.0	20.0	0.0
6052	Cazulas	6	50.0	33.3	0.0	16.7
6073	Barbara	5	40.0	0.0	0.0	60.0

could also have considerable influence on the flow responses, due to the lower connectivity expected between runoff sink and source areas and to the effects of spatial variability of convective summer storms in large basins (for instance in Barbara and Chono) or in basins where the orographic configuration influences the distribution of rainfall (basins further to the East).

Finally, soil moisture content is determined not only by the climate, but also by human activity, so as the PCA results indicate, the basins situated in the first and second quadrants tend to have larger areas of rangeland with typical Mediterranean woodland. The effects of soil use can explain the differences in soil runoff and erosion within the same eco-geomorphological area (Cerdá, 1997). A good example of this could be the experiments to simulate rainfall in limestone areas of SE Spain where Calvo-Cases et al. (2003) found positive correlations between vegetation cover and infiltration rates under dry and wet conditions (summer and winter, respectively). According to these conclusions, a larger surface of rangeland may mean a more homogeneous basin response derived from the impact of the vegetation, improving the soil moisture conditions and the flow connectivity on hillsides and contributing to justifying the IPF linear behaviour.

5. Conclusions

This study proposes an equation to be used on a regional level for the Southeast Basin of Spain to estimate instantaneous peak flows from mean daily flows, drainage area and mean annual rainfall. The expression proposed is based on the linear IPF–MDF behaviour observed in several basins and the application of PCA to extract the most important attributes it displays. It improves the accuracy of the commonly-used Fuller's equation. This suggests that it is a useful equation for estimating maximum peak flows for flood studies and design of hydraulic structures when only mean daily flows are available. The expression to calculate maximum peak flow has been obtained from basins whose drainage areas were between 39 to 182 km^2 and mean annual rainfall between 608 to 888 mm.

The lack of linear relationships IPF–MDF in some basins was connected to the results of the PCA applied. Two types of events were distinguished in terms of their maximum peak flows, which could be explained by the nature of the episodes of maximum rainfall and soil conditions. Maximums that occur in the summer in some of the eastern basins are convective with very intense short storms on bare soil with low moisture content. In contrast, the western basins in the wetter sector, maximums occurring during the rest of the year, can generally be explained by heavy showers from Atlantic squalls, under prior conditions of high soil moisture and vegetation cover. These characteristics limit the use of the equation in arid areas of the Southeast Basin. We recommend more research into this subject in order to develop better estimators, especially in arid areas where traditional hydrological models are not always applicable.

Acknowledgements

This work is being funded by the CICYT Project AGL2002-03400 ("Integración de Procesos Erosivos e Hidrológicos en Cuencas de la Sierra de Cádiz") of the Ministry of Science and Technology (Spain). The first author was supported by a pre-doctoral grant ("CAO01-001-C4-3) from the Junta de Andalucia (Spain).

References

Agencia Andaluza del Agua, 2005. Cuenca Mediterránea Andaluza. Ámbito territorial de la Cuenca Sur. Consejería de Medio Ambiente de la Junta de Andalucía. http://www.agenciaandaluzadelagua.com/v2/index.php?s=generalandsub=entornoande=cuencasur.
Álvarez, M., Puertas, J., Soto, B., Díaz, F., 1999. Análisis Regional de las precipitaciones máximas en Galicia mediante el método del índice de avenida. Ingeniería de Aguas 6 (4), 379–386.
Calvo-Cases, A., Boix-Fayos, C., Imenson, A.C., 2003. Runoff generation, sediment movement and soil water behaviour on calcareous (limestone) slopes of some Mediterranean environments in southeast Spain. Geomorphology 50, 269–291.
Cerdá, A., Lavee, H., 1995. Escorrentía y erosión en los suelos del desierto de Judea. Geographicalia 32, 17–36.
Cerdá, A., 1997. Seasonal changes of the infiltration rates in a Mediterranean scrubland on limestone. J. Hydrol. 198, 209–225.
Cerdá, A., 1998. Relationships between climate and soil hydrological and erosional characteristics along climatic gradients in Mediterranean limestone areas. Geomorphology 25, 123–134.
Correia, F.N., 1983. Metodos de analise e determinacao de caudais de cheia, Laboratorio Nacional de Engenheria Civil, Lisboa, Portugal.
ESRI, Environmental Systems Research Instituteor, 2000. Getting to know Arcview GIS: the geographic information system (GIS) for everyone. Redlands, Calif., USA.
Fill, H.D., 1994. Improving flood quantile estimates using regional information. PhD Thesis, Cornell University, Ithaca, NY, USA.
Fill, H.D., Steiner, A.A., 2003. Estimating instantaneous peak flow from mean daily flow data. J. Hydrol. Eng, ASCE 8 (6), 365–369.
Fuller, W.E., 1914. Flood flows. Trans. Am. Soc. Civ. Eng. 77, 564–617.
Gray, D.M., 1973. Handbook of the Principles of Hydrology. Water information Center, Hungtington, NY, USA.
Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis. Cambridge University Press, Cambridge.
ICA- Instituto Cartografico de Andalucia, 1999. Elevation Model of Andalucía. Junta de Andalucia. Sevilla. Spain.
ICA- Instituto Cartografico de Andalucia, 2000. Interactive Atlas of Andalusia. Junta de Andalucia. Sevilla. Spain.
Kosmas, C., Danalatos, N., Cammeraat, L.H., Chabart, M., Diamantopoulus, J., Farand, R., Gutiérrez, L., Jacob, A., Marques, H., Martínez-Fernández, J., Mizara, A., Moustakas, N., Nicolau, J.M., Oliveiros, C., Pinna, G., Puddu, R., Puigdefábregas, J., Roxo, M., Simao, A., Stamou, G., Tomasi, N., Usai, D., Vacca, A., 1997. The effect of land use on runoff and soil erosion rates under Mediterranean conditions. Catena 29, 45–59.
Landsberg, H.E., 1970. World Survey of Climatology. Climates of Northern and Western Europe, vol. 5. Elsevier, Amsterdam.
Leopold, L.B., Wolman, M.G., Miller, J.P., 1964. Fluvial Processes in Geomorphology. Wiley, San Francisco.
Malinowski, E.R., 1991. Factor Analysis in Chemistry, 2ª Ed. John Wiley, New York.
Mitchel, D.J., 1990. The use of vegetation and land use parameters in modelling catchment sediment yields. In: Thornes, J.B. (Ed.), Vegetation and Erosion. Processes and Environments. Wiley, Chinchester, pp. 289–314.
Mulligan, M., 1998. Modelling the geomorphological impact of climatic variability and extreme events in a semi-arid environment. Geomorphology 24, 59–78.
Nicolau, J.M., Solé-Benet, A., Puigdefábregas, J., Gutiérrez, L., 1996. Effects of soil and vegetation on runoff along a catena in semi-arid Spain. Geomorphology 14, 297–309.
Philippeau, G., 1986. STAT-ITCF: Comment interpréter. les résultats d'une analyse en composantes principales. Services des Études Statistiques, Paris.
Puigdefabregas, J., Del Barrio, G., Boer, M., Gutiérrez, L., Solé, A., 1998. Differential responses of hillslope and channel elements to rainfall events in a semi-arid area. Geomorphology 23, 337–351.
Silva, E.A., Tucci, C.E., 1998. Relacao entre as vazoes maximas diarias e instantaneas. Revista Brasileira de Recuersos Hidricos 3 (1), 133–151.
Stat Soft Incorporation, 2001. Statistica 6.0.Tulsa, Oklahoma, USA.
Stedinger, J.R., 1980. Fitting log-normal distributions to hydrologic data. Water Resour. Res. 16 (3), 481–490.
Steel, R., Torrie, J., 1980. Principles and Procedures of Statistics. A Biometrical Approach, (2 Ed.). Mc Graw Hill, New York.
Tucci, C.E., 1991. Regionalizacao de vazoes no Rio Grande do Sul, IPH/UFRGS, Porto Alegre, Brazil.
Wigley, T.M.L., 1992. Future climate of the Mediterranean Basin with particular emphasis on changes in precipitation. In: Jeftic, L., Millman, J.D., Sestini, G. (Eds.), Climate Change and the Mediterranean Region. Edward Arnold, London, pp. 15–44.